INTRODUCTION TO STATISTICAL MACHINE LEARNING

Statistics 413 – Fall 2019

SYLLABUS

Instructor
Giuseppe Vinci
Office: Duncan Hall 2060
Office Hour: Thursdays 5:15 - 6:15pm
Email: giuseppe.vinci@rice.edu

TA
Zhenwei Dai
Office: Duncan Hall 1036
Office Hour: Tuesdays 6:00 - 7:00pm
Email: zhenwei.dai@rice.edu

TA
Hoai Nam Nguyen
Office: Duncan Hall 2090
Office Hour: Wednesdays 8:00 - 9:00pm
Email: hn17@rice.edu

Times & Locations:

Class: Tuesdays & Thursdays 4:00 - 5:15pm	Location: Herzstein Hall (HRZ) room 210
Recitation: Wednesdays 7:00 - 8:00pm	Location: Herzstein Hall (HRZ) room 210

Course Description: This course is an introduction to concepts, methods, and best practices in statistical machine learning. Topics covered include regularized regression, classification, kernels, dimension reduction, clustering, trees, and ensemble learning. Emphasis will be placed on applied data analysis.

Course Objectives & Outcomes: Students will learn how and when to apply statistical learning techniques, their comparative strengths and weaknesses, and how to critically evaluate the performance of learning algorithms. Students completing this course should be able to (i) apply basic statistical learning methods to build predictive models or perform exploratory analysis, (ii) properly tune, select and validate statistical learning models, and (iii) interpret their results.

Prerequisites: Scientific programming in a language such as R (recommended), Matlab, or Python. Courses: Stat 405, Caam 210, Comp 140, or Comp 130, and linear regression Stat 410. Recommended prerequisites: Matrix analysis or linear algebra (Caam 335 or Math 355) and advanced statistical methods (Stat 411). LaTeX typesetting.

Textbooks:

⋆ *An Introduction to Statistical Learning*, by James, Witten, Tibshirani & Hastie (also available online).

Other books:

- *The Elements of Statistical Learning*, by Hastie, Tibshirani & Friedman (also available online).
- *Statistical Learning with Sparsity*, by Hastie, Tibshirani & Wainwright (also available online).
- *Convex optimization*, by Boyd and Vandenberghe (also available online).

Course Webpages: Canvas and https://gvincistat.wordpress.com/rice-stat413/. Please check Canvas and the course webpage frequently for announcements, and homework and quiz assignments. Please use Canvas to ask questions, share resources, and participate in discussions outside of class.

Grading Policy:

Homeworks	35%
Mid-Term Exam	40%
Quizzes	5%
Project	20%

Attendance Policy: Students are encouraged to attend lectures and recitations.

Homeworks: There will be six homework assignments. Approximate dates:

	Assignment date	Due date
Homework 1	September 8	September 19
Homework 2	September 21	October 3
Homework 3	October 6	October 17
Homework 4	October 20	October 31
Homework 5	November 3	November 14
Homework 6	November 12	November 21

A hard copy of the homework is due at 4:00pm on the due date, and can be turned in at class or to the TA. An electronic copy of the homework must also be submitted on Canvas or in the ways defined in the homework assignment. There will be a deduction of 25% of the grade for homeworks that are not typeset using LaTeX. Late homeworks will NOT be accepted; NO exceptions. Homework topics may be discussed with classmates. You may not, however, share written work or code with others. Plagiarism will NOT be tolerated and will be punished according to the Rice University honor code.

Midterm Exam: There will be an in-class midterm exam assigned on **November 7**. The exam is closed books. However a two-page "cheat sheet" will be allowed with format defined by the instructor. The use of any electronic device is forbidden during the exam. The exam must be completed individually according to the honor code. Plagiarism will NOT be tolerated and will be punished according to the Rice University honor code.

Quizzes: Quizzes are very short sets of questions that will be assigned either as in-class or online problems by the instructor. The purpose of quizzes is to check the level of understanding of specific aspects of the course. About 10 quizzes will be assigned.

Projects: Students are required to work on group projects of statistical learning. Each team may consist of at least two and no more than four students who must equally contribute to their project. Projects will be assigned by the instructor. Team organization will be determined by the instructor. Grades will be based upon two progress reports and a final report (one per team) as well as the results of a final slide presentation. Reports must be typeset in LaTeX. Further details about the projects along with specific grading criteria will be given in a separate document and discussed in class. Approximate dates:

Projects assignment	September 17
Progress Report I	October 15
Progress Report II	November 12
Projects Presentations	December 3 and 5
Final Report Due	December 6

Class schedule will be provided on course websites.

Students with a documented disability requiring accommodations should speak with the instructor during the first two weeks of class.

NOTE: This syllabus is subject to change with reasonable advance notice by the instructor.